

NATIONAL BUREAU OF STANDARDS REPORT

8677

An Introduction to the Statistical Treatment of Measurement Data



U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

THE NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. Its responsibilities include development and maintenance of the national standards of measurement, and the provisions of means for making measurements consistent with those standards; determination of physical constants and properties of materials; development of methods for testing materials, mechanisms, and structures, and making such tests as may be necessary, particularly for government agencies; cooperation in the establishment of standard practices for incorporation in codes and specifications; advisory service to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; assistance to industry, business, and consumers in the development and acceptance of commercial standards and simplified trade practice recommendations; administration of programs in cooperation with United States business groups and standards organizations for the development of international standards of practice; and maintenance of a clearinghouse for the collection and dissemination of scientific, technical, and engineering information. The scope of the Bureau's activities is suggested in the following listing of its four Institutes and their organizational units.

Institute for Basic Standards. Applied Mathematics. Electricity. Metrology. Mechanics. Heat. Atomic Physics. Physical Chemistry. Laboratory Astrophysics.* Radiation Physics. Radio Standards Laboratory.* Radio Standards Physics; Radio Standards Engineering. Office of Standard Reference Data.

Institute for Materials Research. Analytical Chemistry. Polymers. Metallurgy. Inorganic Materials. Reactor Radiations. Cryogenics.* Materials Evaluation Laboratory. Office of Standard Reference Materials.

Institute for Applied Technology. Building Research. Information Technology. Performance Test Development. Electronic Instrumentation. Textile and Apparel Technology Center. Technical Analysis. Office of Weights and Measures. Office of Engineering Standards. Office of Invention and Innovation. Office of Technical Resources. Clearinghouse for Federal Scientific and Technical Information.**

Central Radio Propagation Laboratory.* Ionospheric Telecommunications. Tropospheric Telecommunications. Space Environment Forecasting. Aeronomy.

* Located at Boulder, Colorado 80301.

** Located at 5285 Port Royal Road, Springfield, Virginia 22171.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT

20503-12-2050131

NBS REPORT

8677

April 19, 1965

An Introduction to the Statistical Treatment of Measurement Data

by

Harry H. Ku

IMPORTANT NOTICE

NATIONAL BUREAU OF STANDARDS
for use within the Government. Be
and review. For this reason, the p
whole or in part, is not authorize
Bureau of Standards, Washington, D
the Report has been specifically pre

Approved for public release by the
director of the National Institute of
Standards and Technology (NIST)
on October 9, 2015

accounting documents intended
jected to additional evaluation
sting of this Report, either in
Office of the Director, National
the Government agency for which
ies for its own use.



U.S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

PREFACE

The material contained in this report was prepared originally as two chapters of a textbook on metrology. It deals with basic statistical concepts as related to a measurement process and gives certain selected statistical techniques for the analysis of measurement data. Its sole aim is to introduce to metrologists and physical scientists some of the possible applications of statistical methodology in the field of measurement science, and to do so in a minimum number of pages.

Beginning with the differentiation between arithmetic and measurement numbers, the properties of the latter are developed and described, leading to a discussion of precision and accuracy at the end of the first chapter.

A basic kit of tools for the comparison and manipulation of means and variances is given in the second chapter, including a collection of propagation of error formulas. The use of control chart techniques for monitoring stability is emphasized. Examples are given using actual calibration data of the Bureau. Selected references are given for topics introduced but not treated in detail.

Encouragement and helpful comments from all the members of the Statistical Engineering Laboratory are gratefully acknowledged.

Harry H. Ku

An Introduction to the Statistical Treatment of Measurement Data

<u>Chapter I Statistical Concepts of a Measurement Process</u>	<u>Page</u>
1. Arithmetic numbers and measurement numbers	1
(a) Computation and reporting of results	2
2. Properties of Measurement numbers	3
(a) The arithmetic mean	3
(b) Range, variance and standard deviation	4
(c) Population and frequency curve	5
(d) The normal distribution	7
(e) Estimation of population characteristics	11
3. Interpretation and computation of confidence intervals and limits	11
4. Precision and Accuracy	15
References	18
<u>Chapter II Statistical Analysis of Measurement Data</u>	
1. Summary of Statistical Concepts of a Measurement Process	19
2. Algebra for the manipulation of limiting means and variances	20
(a) Basic Formulas	20
(b) Propagation of error formulas	23
(c) Pooling estimates of variances	26
(d) Components of variances between groups	27
3. Comparison of means and variances	28
(a) Comparison of mean with a standard value	29
(b) Comparison between two or more means	31
(c) Comparison of variances or ranges	34
4. Control chart technique for maintaining stability and precision	35
5. Linear relationships and fitting of constants by the method of least squares	42
References	44

AN INTRODUCTION TO THE STATISTICAL TREATMENT OF MEASUREMENT DATA

Harry H. Ku

Chapter I. Statistical Concepts of a Measurement Process

1. Arithmetic numbers and measurement numbers

In metrological work, digital numbers are used for different purposes and consequently these numbers have different interpretations. It is therefore important to differentiate the two types of numbers which will be encountered.

Arithmetic numbers are exact numbers. 3, $\sqrt{2}$, 1/3, e, or π are all exact numbers by definition, although in expressing some of these numbers in digital form, approximation may have to be used. Thus, π may be written as 3.14 or 3.1416, depending on our judgement of which is the proper one to use from the combined point of view of accuracy and convenience. By the usual rules of rounding, the approximations do not differ from the exact values by more than ± 0.5 units of the last recorded digit. The accuracy of the result can always be extended if necessary.

Measurement numbers, on the other hand, are not approximations to exact numbers, but numbers obtained by operation under approximately the same conditions. For example, three measurements on the diameter of a steel shaft with a micrometer may yield the following results:

<u>no.</u>	<u>diameter in cm.</u>	<u>general notation</u>
1	.396	x_1
2	.392	x_2
3	.401	x_3
	<hr/>	<hr/>
	Sum 1.189	$\sum_{i=1}^n x_i$
		$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
	Average .3963	
	Range .009	$R = x_{\max.} - x_{\min.}$

There is no rounding off here. The last digit in the measured value depends on the instrument used and our ability to read it. If we had used a coarser instrument, we might have obtained 0.4, 0.4, and 0.4; if a finer instrument, we might be able to record to the fifth digit after the decimal point. In all cases, however, the last digit given certainly does not imply that the measured value differs from the diameter D by less than ± 0.5 unit of the last digit.

Thus we see that measurement numbers differ by their very nature from arithmetic numbers. In fact, the phrase "significant figures" has little meaning in the manipulation of numbers resulting from measurements. Reflection on the simple example above will help to convince one of this fact.

(a) Computation and Reporting of Results

By experience, the metrologist can usually select an instrument to give him results adequate for his needs, as illustrated in the example above. Unfortunately, in the process of computation, both arithmetic numbers and measurement numbers are present, and frequently confusion reigns over the number of digits to be kept in successive arithmetic operations.

No general rule can be given for all types of arithmetic operations. If the instrument is well-chosen, severe rounding would result in loss of information. One suggestion, therefore, is to treat all measurement numbers as exact numbers in the operations and to round off the final result only. Another recommended procedure is to carry two or three extra figures throughout the computation, and then to round off the final reported value to an appropriate number of digits.

The "appropriate" number of digits to be retained in the final result depends on the "uncertainties" attached to this reported value. The term "uncertainty" will be treated later under Precision and Accuracy, our only concern here is the number of digits in the expression for uncertainty.

A recommended rule is that the uncertainty should be stated to no more than two significant figures, and the reported value itself should be stated to the last place affected by the qualification given by the uncertainty statement. An example is:

"The apparent mass correction for the nominal 10 g. weight is +0.0420 mg. with a overall uncertainty of ± 0.0087 mg. using three standard deviations as a limit to the effect of random errors of measurement, the magnitude of systematic errors from known sources being negligible."

The sentence form is preferred since then the burden is on the reporter to specify exactly the meaning of the term uncertainty, and to spell out its components. Abbreviated forms such as $a \pm b$, where "a" is the reported value and "b" a measure of uncertainty in some vague sense, should always be avoided.

2. Properties of Measurement Numbers

The study of the properties of measurement numbers, or the Theory of Errors, formally began with Thomas Simpson more than 200 years ago, and attained its full development in the hands of Laplace and Gauss. In the next sections some of the important properties of measurement numbers will be discussed and summarized, thus providing a basis for the statistical treatment and analysis of these numbers in the following chapter.

(a) The Limiting Mean

As shown in the micrometer example above, the results of repeated measurements of a single physical quantity under essentially the same conditions yield a set of measurement numbers. Each member of this set is an estimate of the quantity being measured, and has equal claims on its value. By convention, the numerical values of these n measurements are denoted by x_1, x_2, \dots, x_n , the arithmetic mean by \bar{x} , and the range by R , i.e., the difference between the largest value and the smallest value obtained in the n measurements.

If the results of measurements are to make any sense for the purpose at hand, we must require these numbers, though different, to behave as a group in a certain predictable manner. Experience has shown that this is indeed the case under the conditions stated in italics above. In fact, let us adopt as the Postulate of Measurement [2] a statement due to N. Ernest Dorsey:

"The mean of a family of measurements - of a number of measurements for a given quantity carried out by the same apparatus, procedure and observer - approaches a definite value as the number of measurements is indefinitely increased. Otherwise, they could not properly be called measurements of a given quantity. In the theory of errors, this limiting mean is frequently called the 'true' value, although it bears no necessary relation to the true quaesitum, to the actual value of the quantity that the observer desires to measure. This has often confused the unwary. Let us call it the limiting mean."

Thus, according to this postulate, there exists a limiting mean "m" to which \bar{x} approaches as the number of measurements increases indefinitely, or, in symbols $\bar{x} \rightarrow m$ as $n \rightarrow \infty$. Furthermore, if the true value is " τ ", there is usually a difference between m and τ , or $\Delta = m - \tau$, where Δ is defined as the bias or the systematic error of the measurements.

In practice, however, we will run into difficulties. The value of m cannot be obtained since one cannot make an indefinite number of measurements. Even for a large number of measurements, the conditions will not remain constant since changes occur from hour to hour, and from day to day. The value of τ is unknown and usually unknowable, hence also the bias. Nevertheless, this seemingly simple postulate does provide a sound foundation to build on toward a mathematical model, from which estimates can be made and inference drawn, as we shall see later on.

(b) Range, Variance and Standard Deviation

The range of n measurements, on the other hand, does not enjoy this desirable property of the arithmetic mean. With one more measurement, the range may increase but cannot decrease. Since only the largest and the smallest numbers enter into its calculation, obviously the additional information provided by the measurements in between is lost. It will be desirable to look for another measure of the dispersion (spread, or scattering) of our measurements which will utilize each measurement made with equal weight, and which will approach a definite number as the number of measurements is indefinitely increased.

A number of such measures can be constructed, the most frequently used are the variance and the standard deviation. The choice of the variance as the measure of dispersion is based upon its mathematical convenience and maneuverability. Variance is defined as the value approached by the average of the sum of squares of the deviations of individual measurement from the limiting mean as the number of measurements is indefinitely increased, or in symbols:

$$\frac{1}{n} \sum (x_i - m)^2 \rightarrow \sigma^2 = \text{Variance, as } n \rightarrow \infty.$$

The positive square root of the variance, σ , is called the standard deviation (of a single measurement); the standard deviation is of the same dimensionality as the limiting mean.

There are other measures of dispersion such as average deviation and probable error. The relationships between these measures and the standard deviation can be found in reference [1].

(c) Population and the Frequency Curve

We shall call the limiting mean m the location parameter and the standard deviation σ the scale parameter of the population of measurement numbers generated by a particular measurement process. By population is meant the conceptually infinite number of measurements that can be generated. The two numbers m and σ describe this population of measurements to a large extent, and specify it completely in one important special case.

Our model of a measurement process consists then of a defined population of measurement numbers with a limiting mean m and a standard deviation σ . The result of a single measurement, ^{1/} X , can take randomly any of the values belonging to this population. The probability that a particular measurement yields a value of X which is less than or equal to x' is the proportion of the population that is less than or equal to x' , in symbols

$$P\{X \leq x'\} = \text{proportion of population less than or equal to } x'.$$

Similar statements can be made for the probability that X will be greater than or equal to x'' , or for X between x' and x'' as follows:

$$P\{X \geq x''\}$$

$$P\{x' \leq X \leq x''\}$$

For a measurement process that yields numbers on a continuous scale, the distribution of values of X for the population can be represented by a smooth curve, for example, curve C in figure 1. C is called a frequency curve. The area between C and the abscissa bounded by any two values (x_1 and x_2) is the proportion of the population that takes values between the two values, or the probability that X will assume values between x_1 and x_2 . For example, the probability that $X \leq x'$, can be represented by the shaded area to the left of x' ; the total area between the frequency curve and the abscissa being one by definition.

^{1/} We shall follow here the convention in using the capital X to represent the value that might be produced by employing the measurement process to obtain a measurement (i.e., a random variable), and the lower case x to represent a particular value of X observed.

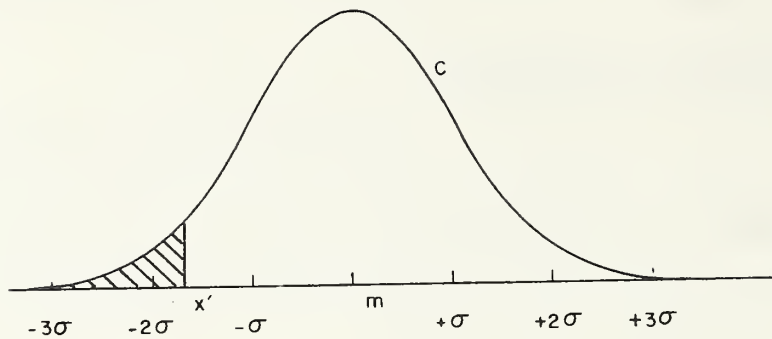


Figure 1 A symmetrical distribution.

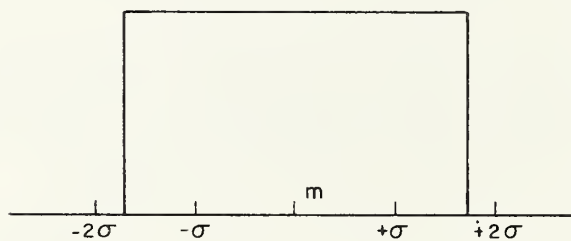
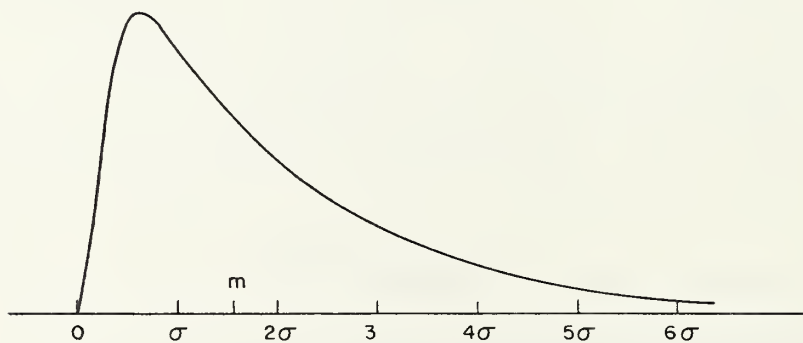


Figure 2a The uniform distribution.



2b The log-normal distribution.

Note that the shape of C is not determined by m and σ alone. Any curve C' enclosing an area of unity with the abscissa defines the distribution of a particular population. Two examples, the uniform distribution and the log-normal distribution are given in figures 2a and 2b. These and other distributions are useful in describing certain populations.

(d) The Normal Distribution

For data generated by a measurement process, the following properties are usually observed:

- i. The results spread roughly symmetrically about a central value.
- ii. Small deviations from this central value are more frequently found than large deviations.

A measurement process having these two properties would generate a frequency curve similar to that shown in figure 1 which is symmetrical and bunched together about m . The study of a particular theoretical representation of a frequency curve of this type leads to the celebrated bell-shaped normal curve (Gauss error curve). Measurements having such a normal frequency curve are said to be normally distributed, or distributed in accordance with the normal law of error.

The normal curve can be represented exactly by the mathematical expression:

$$y = \frac{1}{\sqrt{2\pi} \sigma} \exp \left[-\frac{1}{2} \frac{(x-m)^2}{\sigma^2} \right] \quad (1.1)$$

where y is the ordinate and x the abscissa and $e \doteq 2.71828$ is the base of natural logarithms.

Some of the important features of the normal curve are:

- (i) It is symmetrical about m .
- (ii) The area under the curve is one, as required.
- (iii) If σ is used as unit on the abscissa, then the area under the curve between constant multiples of σ can be computed from tabulated values of the normal distribution. In particular, areas under the curve for some useful intervals between $m - k\sigma$ and $m + k\sigma$ are given in table 1.

Table 1

Area under normal curve between $m - k\sigma$ and $m + k\sigma$

k:	.6745	1.00	1.96	2.00	2.58	3.00
Percent area under curve: (approx.)	50.0	68.4	95.0	95.5	99.0	99.7

Thus about two thirds of the area lies within one σ of m , more than 95 percent within 2σ of m , and less than 0.3 percent beyond 3σ from m .

- (iv) From expression (1.1), it is evident that the frequency curve is completely determined by the two parameters m and σ .

The normal distribution has been studied intensively during the past century. Consequently, if the measurements follow a normal distribution, we can say a great deal about the measurement process. The question remains: How do we know that this is so from the limited number of repeated measurements on hand?

The answer is that we don't! However, in most instances the metrologist may be willing:

- (A) To assume that the measurement process generates numbers that follow a normal distribution approximately, and act as if this were so,
- (B) To rely on the so-called Central Limit Theorem, one version of which is the following^{1/}: "If a population has a finite variance σ^2 and mean m , then the distribution of the sample mean (of n independent measurements) approaches the normal distribution with variance σ^2/n and mean m as the sample size n increases." This remarkable and powerful theorem is indeed tailored for measurement processes. First, every measurement process must by definition have a finite mean and variance. Second, the sample mean \bar{x} is the quantity of interest which, according to the theorem, will be approximately normally distributed for large sample sizes. Third, the measure of dispersion, i.e., the standard

^{1/} From Chapter 7, Introduction to the Theory of Statistics, by A.M. Mood, McGraw-Hill, 1950.

deviation of the sample mean, is reduced by a factor of $1/\sqrt{n}$! This last statement is true in general for all measurement processes in which the measurements are "independent" and for all n . It is therefore not a consequence of the Central Limit Theorem. The Theorem guarantees, however, that the distribution of sample means of independent measurements will be approximately normal with the specified limiting mean and standard deviation σ/\sqrt{n} for large n .

In fact, for a measurement process with a frequency curve that is symmetrical about the mean and with small deviations from the mean as compared to the magnitude of the quantity measured, the normal approximation to the distribution of \bar{x} becomes very good even for n as small as 3 or 4. Fig. 3 shows the uniform and normal distribution having the same mean and standard deviation. The peaked curve is actually two curves, representing the distribution of arithmetic means of four independent measurements from the respective distributions. These curves are indistinguishable to this scale.

A formal definition of the concept of "independence" is out of the scope here. Intuitively, we may say that n normally distributed measurements are independent if these measurements are not correlated or associated in any way. Thus, a sequence of measurements showing a trend or pattern are not independent measurements.

There are many ways by which dependence or correlation creeps into a set of measurement data; several of the common causes are the following:

- (i) Measurements are correlated through a factor that has not been considered, or has been considered to be of no appreciable effect on the results.
- (ii) A standard correction constant has been used for a factor, e.g., temperature, but the constant may overcorrect or undercorrect for particular samples.
- (iii) Measurements are correlated through time of the day, between days, weeks, or seasons.
- (iv) Measurements are correlated through rejection of valid data, when the rejection is based on the size of the number in relation to others of the group.

The traditional way of plotting the data in the sequence they are taken, or in some rational grouping, is perhaps still the most effective way of detecting trends or correlation.

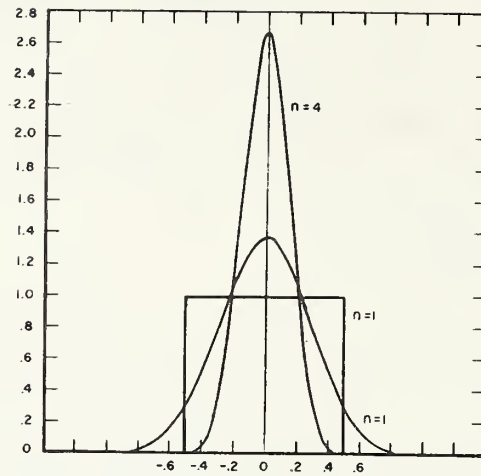


Figure 3 Uniform and normal distribution of individual measurements having the same mean and standard deviation, and the corresponding distribution(s) of arithmetic means of four independent measurements.

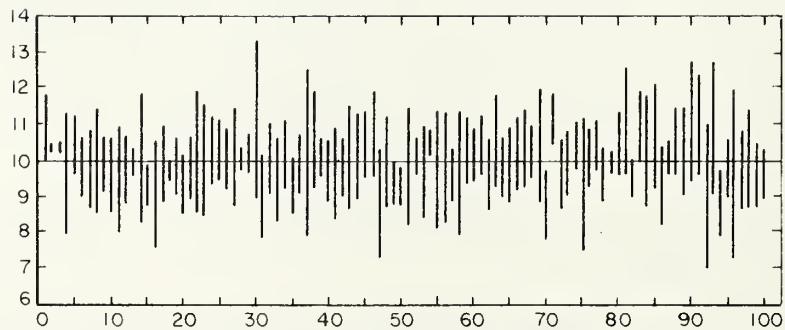


Figure 4 Computed 90% confidence intervals for 100 samples of size 4 drawn at random from a normal population with $m = 10$, $\sigma = 1$.

(e) Estimates of Population Characteristics

In the above section it is shown that the limiting mean, m , and the variance, σ^2 , completely specify a measurement process that follows the normal distribution. In practice, m and σ^2 are not known and cannot be computed from a finite number of measurements. This leads to the use of the sample mean, \bar{x} , as an estimate of the limiting mean, m , and s^2 , the square of the computed standard deviation of the sample, as an estimate of the variance. The standard deviation of the average of n measurements, σ/\sqrt{n} , is sometimes referred to as the standard error of the mean, and is estimated by s/\sqrt{n} .

We note that the making of n independent measurements is equivalent to drawing a sample of size n at random from the population. Two concepts are of importance here:

First, the measurement process is established and under control, meaning the limiting mean and the standard deviation do possess definite values which will not change over a reasonable period of time.

Secondly, the measurements are randomly drawn from this population, implying that the values are of equal weights, and there is no prejudice in the method of selection. Suppose out of three measurements the one which is far apart from the other two is rejected, then the result will not be a random sample.

For a random sample we can say that \bar{x} is an unbiased estimate of m and s^2 is an unbiased estimate of σ^2 , i.e., the limiting mean of \bar{x} is equal to m ; and of s^2 to σ^2 , where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \text{ and}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left[\sum x_i^2 - \frac{(\sum x_i)^2}{n} \right].$$

In addition, we define $s = \sqrt{s^2}$ = computed standard deviation.

3. Interpretation and Computation of Confidence Interval and Limits

By making k sets of n measurements each we can compute and arrange the k values of \bar{x} and s in tabular form as follows:

<u>Set</u>	<u>Sample Mean</u>	<u>Sample Standard Dev.</u>
1	\bar{x}_1	s_1
2	\bar{x}_2	s_2
.	.	.
.	.	.
.	.	.
j	\bar{x}_j	s_j
.	.	.
.	.	.
k	\bar{x}_k	s_k

In the array of \bar{x} 's no two will be likely to have exactly the same value. From the Central Limit Theorem we deduce that the \bar{x} 's will be approximately normally distributed with standard deviation σ/\sqrt{n} . The frequency curve of \bar{x} will be centered about the limiting mean m and will have the scale factor σ/\sqrt{n} . In other words, $\bar{x} - m$ will be centered about zero, and the quantity

$$z = \frac{\bar{x} - m}{\sigma/\sqrt{n}}$$

has the properties of a single observation from the "standardized" normal distribution which has a mean zero and a standard deviation of one.

From tabulated values of standardized normal distribution we know that 95% of z values will be bounded between -1.96 and $+1.96$. Hence the statement

$$-1.96 \leq \frac{\bar{x} - m}{\sigma/\sqrt{n}} \leq +1.96 ,$$

or its equivalent,

$$\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} ,$$

will be correct 95% of the time in the long run. The interval $\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}$ to $\bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$ is called a confidence interval for m . The probability that the confidence interval will cover the limiting mean, 0.95 in this case, is called the confidence level or confidence coefficient. The values of the end points of a confidence interval are called confidence limits. It is to be borne in mind that \bar{x} will fluctuate from set to set, and the interval calculated for a particular \bar{x}_j may or may not cover m .

In the above discussion we have selected a two-sided interval symmetrical about \bar{x} . For such intervals the confidence coefficient is usually denoted by $1 - \alpha$, where $\alpha/2$ is the proportion of the area under the frequency curve of z that is cut off from each tail.

In most cases, σ is not known and an estimate of σ is computed from the same set of measurements we use to calculate \bar{x} . Nevertheless, let us form a quantity similar to z which is:

$$t = \frac{\bar{x} - m}{s/\sqrt{n}},$$

and if we know the distribution of t , we could make the same type of statement as before. In fact the distribution of t is known for the case of normally distributed measurements.

The distribution of t was obtained mathematically by William S. Gosset under the pen name of "Student", hence the distribution of " t " is called the Student's distribution. In the expression for t , both \bar{x} and s fluctuate from set to set of measurements. Intuitively we will expect the value of t to be larger than that of z for a statement with the same probability of being correct. This is indeed the case. the values of t are listed in table 2.

Table 2^{1/}

A Brief Table of Values of t

<u>degrees of freedom</u>	<u>Confidence Level: $P = 1 - \alpha$</u>			
	<u>.500</u>	<u>.900</u>	<u>.950</u>	<u>.990</u>
<u>v</u>				
1	1.000	6.314	12.706	63.657
2	.816	2.920	4.303	9.925
3	.765	2.353	3.182	5.841
4	.741	2.132	2.776	4.604
5	.727	2.015	2.571	4.032
6	.718	1.943	2.447	3.707
7	.711	1.895	2.365	3.499
10	.700	1.812	2.228	3.169
15	.691	1.753	2.131	2.947
20	.687	1.725	2.086	2.845
30	.683	1.697	2.042	2.750
60	.679	1.671	2.000	2.660
∞	.674	1.645	1.960	2.576

^{1/} Adapted from Biometrika Tables for Statisticians, Vol.I, Edited by E.S. Pearson and H.O. Hartley, 1958, The University Press, Cambridge.

To find a value for t , we need to know the "degrees of freedom" (ν) associated with the computed standard deviation s . Since \bar{x} is calculated from the same n numbers and has a fixed value, the n^{th} value of x_i is completely determined by \bar{x} and the other $(n-1)$ values. Hence the degrees of freedom here are $n-1$.

Having the table for the distribution of t , and using the same reasoning as before, we can make the statements that:

$$\bar{x} - t \frac{s}{\sqrt{n}} \leq m \leq \bar{x} + t \frac{s}{\sqrt{n}}$$

and our statements will be correct $100(1-\alpha)\%$ of the time in the long run. The value of t depends on the degrees of freedom ν and the probability level. From the table, we get for a confidence level of .95, the following lower and upper confidence limits:

ν	$L_l = \bar{x} - t s/\sqrt{n}$	$L_u = \bar{x} + t s/\sqrt{n}$
1	$\bar{x} - 12.706 s/\sqrt{n}$	$\bar{x} + 12.706 s/\sqrt{n}$
2	$\bar{x} - 4.303 s/\sqrt{n}$	$\bar{x} + 4.303 s/\sqrt{n}$
3	$\bar{x} - 3.182 s/\sqrt{n}$	$\bar{x} + 3.182 s/\sqrt{n}$

The value of t for $\nu = \infty$ is 1.96, the same as for the case of known σ . Notice that very little can be said about m with two measurements. However, for n larger than 2, the interval predicted to contain m narrows down steadily, due to both the smaller value of t and the divisor \sqrt{n} .

It is probably worthwhile to emphasize again that each particular confidence interval computed as a result of n measurements will either include m or fail to include m . The probability statement refers to the fact that if we make a long series of sets of n measurements, and if we compute a confidence interval for m from each set by the prescribed method, we would expect 95% of such intervals to include m .

Fig. 4 shows the 90% confidence intervals ($P = .90$) computed from 100 samples of $n = 4$ from a normal population with $m = 10$, and $\sigma = 1$. Three interesting features are to be noted:

- (i) The number of intervals that include m actually turns out to be 90, the expected number.
- (ii) The surprising variation of the sizes of these intervals.
- (iii) The closeness of the mid-points of these intervals to the line for the mean does not seem to be related to the spread. In samples no. 2 and no. 3, the four values must have been very close together, but both of these intervals failed to include the line for the mean.

From the widths of computed confidence intervals, one may get an intuitive feeling whether the number of measurements n is reasonable and sufficient for the purpose on hand. It is true that, even for small n , the confidence intervals will cover the limiting mean with the specified probability, yet the limits may be so far apart as to be of no practical significance. For detecting a specified magnitude of interest, e.g., the difference between two means, the approximate number of measurements required can be solved by equating the half-width of the confidence interval to this difference and solving for n using σ when known, or using s by trial and error if σ is not known. Tables of sample sizes required for certain prescribed conditions are given in reference [1] of the next chapter.

4. Precision and Accuracy

Index of Precision

Since σ is a measure of the spread of the frequency curve about the limiting mean, we can speak of σ as an index of precision. Thus a measurement process with a standard deviation σ_1 is said to be more precise than another with standard deviation σ_2 if σ_1 is smaller than σ_2 . (In fact, σ is really a measure of imprecision since the imprecision is directly proportional to σ .)

Consider the means of sets of n independent measurements as a new derived measurement process. The standard deviation of the new process is σ/\sqrt{n} . It is therefore possible to derive from a less precise measurement process a new process which has a standard deviation equal to that of a more precise process. This is accomplished by making more measurements.

Suppose $m_1 = m_2$, but $\sigma_1 = 2\sigma_2$. Then for a derived process to have $\sigma'_1 = \sigma_2$, we need

$$\sigma'_1 = \frac{\sigma_1}{\sqrt{n}} = \frac{2\sigma_2}{\sqrt{4}},$$

or we need to use the average of four measurements as a single measurement. Thus for a required degree of precision, the number of measurements, n_1 and n_2 , needed for measurement processes I and II is proportional to the squares of their respective standard deviations (variances), or in symbols

$$\frac{n_1}{n_2} = \frac{\sigma_1^2}{\sigma_2^2}.$$

If σ is not known, and the best estimate we have of σ is a computed standard deviation s based on n measurements then s could be used as an estimate of the index of precision. The value of s , however, may vary considerably from sample to sample in the case of a small number of measurements, as was shown in figure 4, where the lengths of the intervals are constant multiples of s computed from the samples. The number n or the degrees of freedom ν must be considered along with s in indicating how reliable an estimate s is of σ . In what follows, whenever the terms standard deviation about the limiting mean (σ) or standard error of the mean ($\sigma_{\bar{x}}$) are used, the respective estimates s and s/\sqrt{n} may be substituted, by taking into consideration the above reservation.

In metrology or calibration work, the precision of the reported value is an integral part of the result. In fact, precision is the main criterion by which the quality of the work is judged. Hence, the laboratory reporting the value must be prepared to give evidence of the precision claimed. Obviously an estimate of the standard deviation of the measurement process based only on a small number of measurements cannot be considered as convincing evidence. By the use of the control chart method for standard deviation and by the calibration of one's own standard at frequent intervals, as described in sections 4 and 3(a) of the next chapter, the laboratory may eventually claim that the standard deviation is in fact known and the measurement process is stable, with readily available evidence to support these claims.

Interpretation of Precision

Since a measurement process generates numbers as the results of repeated measurements of a single physical quantity under essentially the same conditions, the method and procedure in obtaining these numbers must be specified in detail. However, no amount of detail would cover all the contingencies that may arise, or cover all the factors that may affect the results of measurement. Thus a single operator in a single day with a single instrument may generate a process with a precision index measured by σ . Many operators measuring the same quantity over a period of time with a number of instruments will yield a precision index measured by σ' . Logically σ' must be larger than σ , and in practice it is usually considerably larger. Consequently modifiers of the words "precision" are recommended by ASTM^{1/} to qualify

^{1/} Use of the Terms Precision and Accuracy as Applied to the Measurement of a .
Property of a Material, ASTM Designation: E177-61T, 1961.

in an unambiguous manner what is meant. Examples are "single-operator-machine", "multilaboratory", "single-operator-day" etc. The same publication warns against the use of the terms "repeatability" and "reproducibility" if the interpretation of these terms is not clear from the context.

The standard deviation σ , or the standard error σ/\sqrt{n} can be considered as a yardstick, with which we can gauge the difference between two results obtained as measurements of the same physical quantity. If our interest is to compare the results of one operator against another, the single-operator precision is probably appropriate, and if the two results differ by an amount considered to be large as measured by the standard errors, we may conclude that the evidence is predominantly against the two results being truly equal. In comparing the results of two laboratories, the single-operator precision is obviously an inadequate measure to use, since the precision of each laboratory must include factors such as multi-operator-day-instruments.

Hence the selection of an index of precision depends strongly on the purposes for which the results are to be used or might be used. It is common experience that three measurements made within the hour are closer together than three measurements made on, say, three separate days. However, an index of precision based on the former is generally not a justifiable indicator of the quality of the reported value. For a thorough discussion on the realistic evaluation of precision see section 4 of ref.[2].

Accuracy

The term "accuracy" usually denotes in some sense the closeness of the measured values to the true value, taking into consideration both precision and bias. Bias, defined as the difference between the limiting mean and the true value, is a constant, and does not behave in the same way as the index of precision, the standard deviation. In many instances, the possible sources of biases are known but their magnitudes and directions are not known. The overall bias is of necessity reported in terms of estimated bounds that reasonably include the combined effect of all the elemental biases. Since there are no accepted ways to estimate bounds for elemental biases, or to combine them, these should be reported and discussed in sufficient detail so as to enable others to use their own judgment on the matter.

It is recommended that an index of accuracy be expressed as a pair of numbers, one the credible bounds for bias, and the other an index of precision, usually in the form of a multiple of the standard deviation (or estimated standard deviation). The terms "uncertainty" and "limits of error" are sometimes used to express the sum of these two components, and their meanings are ambiguous unless the components are spelled out in detail.

References

The following references are recommended for topics introduced in this section:

1. An Introduction to Scientific Research, by E. B. Wilson, Jr., McGraw-Hill Book Co., Inc., New York, N.Y., 1952, Chapters 7, 8, and 9.
2. "Realistic evaluation of the precision and accuracy of instrument calibration system" by Churchill Eisenhart, Journal of Research of the National Bureau of Standards, vol.67C, No.2, 1963.
3. Experimentation and Measurement, by W. J. Youden, National Science Teachers Association Vistas of Science Series No.2, Scholastic Book Services, New York 36, N.Y.

Chapter II. Statistical Analysis of Measurement Data

1. Introduction

In the last chapter the basic concepts of a measurement process were given in an expository manner. These concepts, necessary to the statistical analysis to be presented in this chapter, are summarized and reviewed below: By making a measurement we obtain a number intended to express quantitatively a measure of "the property of a thing." Measurement numbers differ from ordinary arithmetic numbers, and the usual "significant figure" treatment is not appropriate. Repeated measurement of a single physical quantity under essentially the same conditions generates a sequence of numbers x_1, x_2, \dots, x_n . A measurement process is established if this conceptually infinite sequence has a limiting mean m and a standard deviation σ .

For many measurement processes encountered in metrology, the sequence of numbers generated follows approximately the normal distribution, specified completely by the two quantities m and σ . Moreover, averages of n independent measurement numbers tend to be normally distributed with the limiting mean m and the standard deviation σ/\sqrt{n} , regardless of the distribution of the original numbers. Normally distributed measurements are independent if they are not correlated or associated in any way. A sequence of measurements showing a trend or pattern are not independent measurements.

Since m and σ are usually not known, these quantities are estimated by calculating \bar{x} and s from n measurements where,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \text{ and} \quad (1.1)$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n} \right]} \quad (1.2)$$

The distribution of the quantity $t = \frac{\bar{x} - m}{s/\sqrt{n}}$ (for x normally distributed) is known. From the tabulated values of t (see p. 13), confidence intervals can be constructed to bracket m for a given confidence coefficient $1 - \alpha$ (probability of being correct in the long run).

The confidence limits are the end points of confidence intervals defined by

$$\begin{aligned} L_l &= \bar{x} - t \frac{s}{\sqrt{n}} \\ L_u &= \bar{x} + t \frac{s}{\sqrt{n}} \end{aligned} \quad (1.3)$$

where the value of t is determined by two parameters, namely, the degrees of freedom v associated with s and the confidence coefficient $1 - \alpha$.

The width of a confidence interval gives an intuitive measure of the uncertainty of the evidence given by the data. Too wide an interval may merely indicate that more measurements need to be made for the objective desired.

2. Algebra for the manipulation of limiting means and variances

(a) Basic Formulas

A number of basic formulas are extremely useful in dealing with a quantity which is a combination of other measured quantities.

- (i) Let m_x and m_y be the respective limiting means of two measured quantities X and Y , and a, b be constants, then

$$\begin{aligned} m_{x+y} &= m_x + m_y, \\ m_{x-y} &= m_x - m_y, \text{ and} \\ m_{ax+by} &= am_x + bm_y. \end{aligned} \quad (2.1)$$

- (ii) If, in addition, X and Y are independent, then it is also true that

$$m_{xy} = m_x m_y. \quad (2.2)$$

For paired values of X and Y , we can form the quantity Z with

$$Z = (X - m_x)(Y - m_y). \quad (2.3)$$

Then by formula (2.2) for independent variables,

$$\begin{aligned} m_Z &= m_{(x-m_x)(y-m_y)} \\ &= (m_x - m_x)(m_y - m_y) = 0. \end{aligned}$$

Thus $m_Z = 0$ when X and Y are independent.

- (iii) The limiting mean of Z in (2.3) is defined as the covariance of X and Y and is usually denoted by $\text{cov}(X, Y)$, or σ_{xy} . The covariance, similar to the variance, is estimated by

$$s_{xy} = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y}) . \quad (2.4)$$

Thus if X and Y are correlated in such a way that paired values are likely to be both higher or lower than their respective means, then s_{xy} tends to be positive. If a high x value is likely to be paired with a low y value, and vice versa, then s_{xy} tends to be negative. If X and Y are not correlated, s_{xy} tends to zero (for large n).

(iv) The correlation coefficient ρ is defined as:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} , \quad (2.5)$$

and is estimated by

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (2.6)$$

Both ρ and r lie between -1 and $+1$.

(v) Let σ_x^2 and σ_y^2 be the respective variances of X and Y , and σ_{xy} the covariance of X and Y , then

$$\begin{aligned} \sigma_{x+y}^2 &= \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy} \\ \sigma_{x-y}^2 &= \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy} \end{aligned} \quad (2.7)$$

If X and Y are independent, $\sigma_{xy} = 0$, then

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2 = \sigma_{x-y}^2 \quad (2.8)$$

Since the variance of a constant is zero, we have

$$\begin{aligned} \sigma_{ax+b}^2 &= a^2 \sigma_x^2 , \text{ and} \\ \sigma_{ax+by}^2 &= a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2ab\sigma_{xy} \end{aligned} \quad (2.9)$$

In particular, if X and Y are independent and normally distributed, then $aX + bY$ is normally distributed with limiting mean $am_x + bm_y$ and variance $a^2\sigma_x^2 + b^2\sigma_y^2$. For measurement situations in general, metrologists usually strive to get measurements that are independent, or can be assumed to be independent. The case when two quantities are dependent because

both are functions of other measured quantities will be treated under propagation of error formulas (see formula (2.13)).

- (vi) Standard errors of the sample mean and the weighted means (of independent measurements) are special cases of the above.

Since $\bar{x} = \frac{1}{n} \sum x_i$ and the x_i 's are independent with variance σ_x^2 , it follows, by (2.9),

$$\begin{aligned}\sigma_{\bar{x}}^2 &= \left(\frac{1}{n}\right)^2 \sigma_{x_1}^2 + \left(\frac{1}{n}\right)^2 \sigma_{x_2}^2 \dots \left(\frac{1}{n}\right)^2 \sigma_{x_n}^2 \\ &= \frac{\sigma_x^2}{n}\end{aligned}\quad (2.10)$$

as previously stated.

If \bar{x}_1 is an average of k values, and \bar{x}_2 is an average of n values, then for the overall average, \bar{x} , it is logical to compute

$$\bar{x} = \frac{x_1 + \dots + x_k + x_{k+1} + \dots + x_{k+n}}{k + n}$$

and $\sigma_{\bar{x}}^2 = \frac{1}{k+n} \sigma_x^2$. However, this is equivalent to a weighted mean of \bar{x}_1 and \bar{x}_2 where the weights are proportional to the number of measurements in each average i.e.,

$$w_1 = k, \quad w_2 = n, \quad \text{and}$$

$$\begin{aligned}\bar{x} &= \left(\frac{w_1}{w_1 + w_2}\right) \bar{x}_1 + \left(\frac{w_2}{w_1 + w_2}\right) \bar{x}_2 \\ &= \frac{k}{n+k} \bar{x}_1 + \frac{n}{n+k} \bar{x}_2.\end{aligned}$$

Since

$$\frac{\sigma_{\bar{x}_1}^2}{\sigma_{\bar{x}_2}^2} = \frac{\sigma^2/k}{\sigma^2/n} = \frac{n}{k} = \frac{w_2}{w_1},$$

the weighting factors w_1 and w_2 are therefore also inversely proportional to the respective variances of the averages. This principle can be extended to more than two variables in the following manner.

Let $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ be a set of averages estimating the same quantity. The overall average may be computed to be

$$\bar{x} = \frac{1}{w_1 + w_2 + \dots + w_k} [w_1 \bar{x}_1 + w_2 \bar{x}_2 + \dots + w_k \bar{x}_k],$$

where

$$w_1 = \frac{1}{\sigma_{\bar{x}_1}^2}, \quad w_2 = \frac{1}{\sigma_{\bar{x}_2}^2}, \quad \dots, \quad w_k = \frac{1}{\sigma_{\bar{x}_k}^2}.$$

The variance of \bar{x} is, by (2.9)

$$\sigma_{\bar{x}}^2 = \frac{1}{w_1 + w_2 + \dots + w_k} . \quad (2.11)$$

In practice, the estimated variances $s_{\bar{x}}^2$ will have to be used in the above formulas, and consequently the equations hold only as approximations.

(b) Propagation of error formulas

The results of a measurement process can usually be expressed by a number of averages $\bar{x}, \bar{y} \dots$, and the standard errors of these averages

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} , \quad s_{\bar{y}} = \frac{s_y}{\sqrt{k}} ,$$

etc. These results, however, may not be of direct interest, whereas the quantity of interest is in the functional relationship $m_w = f(m_x, m_y)$. It is desired to estimate m_w by $\bar{w} = f(\bar{x}, \bar{y})$ and to compute $s_{\bar{w}}$ as an estimate of $\sigma_{\bar{w}}$.

It has been shown that under certain general restrictions, the propagation of error formulas work surprisingly well. The $\sigma_{\bar{x}}^2, \sigma_{\bar{y}}^2$ and $\sigma_{\bar{w}}^2$ that are used in the following formulas will often be replaced in practice by the computed values $s_{\bar{w}}^2, s_{\bar{x}}^2$ and $s_{\bar{y}}^2$.

The general formula for $\sigma_{\bar{w}}^2$ is given by:

$$\sigma_{\bar{w}}^2 = \left[\frac{\partial f}{\partial x} \right]^2 \sigma_{\bar{x}}^2 + \left[\frac{\partial f}{\partial y} \right]^2 \sigma_{\bar{y}}^2 + 2 \left[\frac{\partial f}{\partial x} \right] \left[\frac{\partial f}{\partial y} \right] \rho_{\bar{x}\bar{y}} \sigma_{\bar{x}} \sigma_{\bar{y}} \quad (2.12)$$

where the partial derivatives in square brackets are to be evaluated at the averages of x and y . If X and Y are independent, $\rho = 0$ and therefore the last term equals zero. If X and Y are measured in pairs, $s_{\bar{x}\bar{y}}$ (2.6) can be used as an estimate of $\rho_{\bar{x}\bar{y}} \sigma_{\bar{x}} \sigma_{\bar{y}}$.

If W is functionally related to U and V by

$$m_w = f(m_u, m_v),$$

and both U and V are functionally related to X and Y by

$$m_u = g(m_x, m_y)$$

$$m_v = h(m_x, m_y),$$

then U and V are functionally related. We will need the covariance

$\sigma_{\bar{u}\bar{v}} = \rho_{\bar{u}\bar{v}} \sigma_{\bar{u}} \sigma_{\bar{v}}$ to calculate $\sigma_{\bar{w}}^2$. The covariance $\sigma_{\bar{u}\bar{v}}$ is given approximately by:

$$\sigma_{\bar{u}\bar{v}} = \left[\frac{\partial g}{\partial \bar{x}} \cdot \frac{\partial h}{\partial \bar{x}} \right] \sigma_{\bar{x}}^{-2} + \left[\frac{\partial g}{\partial \bar{y}} \cdot \frac{\partial h}{\partial \bar{y}} \right] \sigma_{\bar{y}}^{-2} \quad (2.13)$$

$$+ \left\{ \left[\frac{\partial g}{\partial \bar{x}} \cdot \frac{\partial h}{\partial \bar{y}} \right] + \left[\frac{\partial g}{\partial \bar{y}} \cdot \frac{\partial h}{\partial \bar{x}} \right] \right\} \rho_{\bar{x}\bar{y}} \sigma_{\bar{x}} \sigma_{\bar{y}}$$

The square brackets mean, as before, that the partial derivatives are to be evaluated at \bar{x} and \bar{y} . If X and Y are independent, the last term again vanishes.

These formulas can be extended to three or more variables if necessary. For convenience, a few special formulas for commonly encountered functions are listed in table 3 with X, Y assumed to be independent. These may be derived from the above formulas as exercises.

Table 3
Propagation of Error Formulas for Some Simple Functions
(X and Y are assumed to be independent)

Function Form	Approx. Formula for s_w^{-2}
$m_w = Am_x + Bm_y$	$A^2 s_{\bar{x}}^{-2} + B^2 s_{\bar{y}}^{-2}$
$m_w = \frac{m_x}{m_y}$	$\left(\frac{\bar{x}}{\bar{y}} \right)^2 \left(\frac{s_{\bar{x}}^{-2}}{\bar{x}^2} + \frac{s_{\bar{y}}^{-2}}{\bar{y}^2} \right)$
$m_w = \frac{1}{m_y}$	$\frac{s_{\bar{y}}^{-2}}{\bar{y}^4}$
$m_w = \frac{m_x}{m_x + m_y}$	$\left(\frac{\bar{w}}{\bar{x}} \right)^4 \left(\frac{s_{\bar{x}}^{-2}}{\bar{y}^2} + \frac{s_{\bar{y}}^{-2}}{\bar{x}^2} \right)$
$m_w = \frac{m_x}{1 + m_x}$	$\frac{s_{\bar{x}}^{-2}}{(1 + \bar{x})^4}$
$*m_w = m_x m_y$	$(\bar{x}\bar{y})^2 \left(\frac{s_{\bar{x}}^{-2}}{\bar{x}^2} + \frac{s_{\bar{y}}^{-2}}{\bar{y}^2} \right)$
$*m_w = m_x^2$	$4\bar{x}^2 s_{\bar{x}}^{-2}$
$m_w = \sqrt{m_x}$	$\frac{1}{4} \frac{s_{\bar{x}}^{-2}}{\bar{x}}$
$*m_w = \ln m_x$	$\frac{s_{\bar{x}}^{-2}}{\bar{x}^2}$
$*m_w = k m_x^a m_y^b$	$\bar{w}^2 \left(a^2 \frac{s_{\bar{x}}^{-2}}{\bar{x}^2} + b^2 \frac{s_{\bar{y}}^{-2}}{\bar{y}^2} \right)$

Function Form

$$*m_w = e^{m_x}$$

$$w = 100 \frac{s_x}{\bar{x}} \quad (= \text{coefficient of variation})$$

Approx. Formula for s_w^{-2}

$$e^{2\bar{x}} s_{\bar{x}}^{-2}$$

$$\frac{w^2}{2(n-1)} \quad (\text{not directly derived from the formulas}) \quad \frac{1}{-}$$

* Distribution of \bar{w} is highly skewed and normal approximation could be seriously in error for small n .

1/ See, for example, Statistical Theory with Engineering Applications, p.301, by A. Hald, John Wiley and Sons, 1952.

In these formulas, if

(a) the partial derivatives when evaluated at the averages are small and

(b) σ_x, σ_y are small compared to \bar{x}, \bar{y} ,

then the approximations are good and \bar{w} tends to be distributed normally (the ones marked by asterisks * are highly skewed and normal approximation could be seriously in error for small n).

c. Pooling estimates of variances

The problem often arises that there are several estimates of a common variance σ^2 which we wish to combine into a single estimate. For example, a gage block may be compared with the master block n_1 times, resulting in an estimate of the variance s_1^2 . Another gage block compared with the master block n_2 times, giving rise to s_2^2 , etc. As long as the nominal thicknesses of these blocks are within a certain range, the precision of calibration can be expected to remain the same. To get a better evaluation of the precision of the calibration process, we would wish to combine these estimates. The rule is to combine the computed variances weighted by their respective degrees of freedom, or

$$s_p^2 = \frac{v_1 s_1^2 + v_2 s_2^2 + \dots + v_k s_k^2}{v_1 + v_2 + \dots + v_k} \quad (2.14)$$

The pooled estimate of the standard deviation, of course, is $\sqrt{s_p^2} = s_p$. In the example $v_1 = n_1 - 1, v_2 = n_2 - 1, \dots, v_k = n_k - 1$, thus the expression reduces to

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_k - 1)s_k^2}{n_1 + n_2 + \dots + n_k - k} \quad (2.15)$$

The degrees of freedom for the pooled estimate is the sum of the degrees of freedom of individual estimates, or $v_1 + v_2 + \dots + v_k = n_1 + n_2 + \dots + n_k - k$. With the increased number of degrees of freedom, s_p is a more dependable estimate of σ than an individual s . Eventually, we may consider the value of s_p to be equal to that of σ and claim that we know the precision of the measuring process.

For the special case where k sets of duplicate measurements are available, the above formula reduces to:

$$s_p^2 = \frac{1}{2k} \sum_{i=1}^k d_i^2 \quad (2.16)$$

where d_i = difference of duplicate readings. The pooled standard deviation s_p has k degrees of freedom.

For sets of normally distributed measurements where the number of measurements in each set is small, say less than ten, an estimate of the standard deviation can be obtained by multiplying the range of these measurements by a constant. Table 4 lists these constants corresponding to the number n of measurements in the set. For large n , considerable information is lost and this procedure is not recommended.

Table 4

Estimate of σ from the range $\frac{1}{n}$

<u>n</u>	<u>Multiplying factor</u>
2	.886
3	.591
4	.486
5	.430
6	.395
7	.370
8	.351
9	.337
10	.325

If there are k sets of n measurements each, the average range \bar{R} can be computed. The standard deviation can be estimated by multiplying the average range by the factor for n .

(d) Components of variance between groups

In pooling estimates of variances from a number of subgroups, we have increased confidence in the value of the estimate obtained. Let us call this estimate the within-group standard deviation, σ_w . The within-group standard deviation σ_w is a proper measure of dispersions of values within the same group, but not necessarily the proper one for dispersions of values belonging to different groups.

If in making calibrations there is a difference between groups, say from day to day, or from set to set, then the limiting means of the groups are not equal. We could think of these limiting means as individual measurements and assume that the

1/ Adapted from Biometrika Tables for Statisticians, vol.1, Edited by E.S. Pearson and H. O. Hartley, 1958, The University Press, Cambridge.

average of these limiting means will approach a limit which we will call the limiting mean for all the groups. In estimating σ_w^2 , the differences of individuals from the respective group means are used. Obviously σ_w does not include the differences between groups. Let us use σ_b^2 to denote the variance corresponding to the differences between groups, i.e., the measure of dispersions of the limiting means of the respective groups about the limiting mean for all groups.

Thus for each individual measurement x , the variance of x has two components, and

$$\sigma^2 = \sigma_b^2 + \sigma_w^2 .$$

For the group mean \bar{x} with n measurements in the group, we have

$$\sigma_{\bar{x}}^2 = \sigma_b^2 + \frac{\sigma_w^2}{n} .$$

If k groups of n measurements are available giving averages $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$, then an estimate of $\sigma_{\bar{x}}^2$ is

$$s_{\bar{x}}^2 = \frac{1}{k-1} \sum_{i=1}^k (\bar{x}_i - \bar{\bar{x}})^2$$

with $k - 1$ degrees of freedom, where $\bar{\bar{x}}$ is the average of all nk measurements.

The resolution of the total variance into components attributable to identifiable causes or factors and the estimation of such components of variance are topics treated under analysis of variance and experimental design. For selected treatments and examples see reference [2], [3], and [5].

3. Comparison of means and variances

Comparison of means is perhaps one of the most frequently used techniques in metrology. The mean obtained from one measurement process may be compared with a standard value; two series of measurements on the same quantity may be compared; or sets of measurement on more than two quantities may be compared to determine homogeneity of the group of means.

It is to be borne in mind that, in all the comparisons discussed below, we are interested in comparing the limiting means. The sample means and the computed standard errors are used to calculate confidence limits on the difference between two means. The "t" statistic derived from normal distribution theory is used in this procedure since we are assuming either the measurement process is normal, or the sample averages are approximately normally distributed.

(a) Comparison of a mean with a standard value

In calibration of class M weights at the National Bureau of Standards, the weights to be calibrated are intercompared with sets of standard weights having "accepted" corrections. Accepted corrections are based on years of experience and considered to be exact to the accuracy required. For instance, the accepted correction for the NB'10 gram weight is $-.4040$ mg.

The NB'10 is treated as an unknown and calibrated with each set of weights tested using an intercomparison scheme based on 100 gm standard weight. Hence the observed correction for NB'10 can be computed for each particular calibration. Table 5 lists eleven observed corrections of NB'10 during May 1963. ^{1/}

Calculated 95% confidence limits from the eleven observed corrections are $-.4041$ and $-.3995$. These values include the accepted value of $-.4040$, and we conclude that the observed corrections agree with the accepted value.

What if the computed confidence limits for the observed correction do not cover the accepted value? Three explanations may be suggested:

1. The accepted value is correct. However, in choosing $\alpha = .05$, we know that 5% of the time in the long run we will make an error in our statement. By chance alone, it is possible that this particular set of limits would not cover the accepted value.
2. The average of the observed corrections does not agree with the accepted value because of certain systematic error, temporary or seasonal, particular to one or several members of this set of data for which no adjustment has been made.
3. The accepted value is incorrect, e.g., the mass of the standard has changed.

In our example, we would be extremely reluctant to agree to the third explanation since we have much more confidence in the accepted value than the value based only on eleven calibrations. We are warned that something may have gone wrong, but not unduly alarmed since such an event will happen purely by chance about once every twenty times.

The control chart for mean with known value, to be discussed in section 4 would be the proper tool to use to monitor the constancy of the correction of the standard mass.

^{1/} Illustrative data supplied by Robert Raybold, Metrology Division, National Bureau of Standards.

Table 5

Computation of Confidence Limits for Observed Corrections, NB' 10 gm.

<u>Date</u>	<u>Observed Correction to std. 10g. wt. in mg.</u>	
5-1-63	-	.4008
"	-	.4053
"	-	.4022
5-2	-	.4075
"	-	.3994
5-3	-	.3986
5-6	-	.4015
"	-	.3992
"	-	.3973
5-7	-	.4071
"	-	.4012
$\Sigma x_i = -4.4201$		$\Sigma x_i^2 = 1.77623417$
$\bar{x} = -.40183\text{mg.}$		$\frac{(\Sigma x_i)^2}{n} = 1.77611673$
		diff. = .00011744

$$s^2 = \frac{1}{n-1} (.00011744) = .000011744$$

$s = .00343$ = computed std. dev. of an observed correction about the mean.

$$\frac{s}{\sqrt{n}} = .00103 = \text{computed std. dev. of the mean of eleven corrections}$$

$$= \text{computed std. error of the mean}$$

For a two-sided 95% confidence interval for the mean of the above sample of size 11, $\alpha/2 = 0.025$, $v = 10$, and the corresponding value of t is equal to 2.228 in the table of "t" distribution. Therefore,

$$L_l = \bar{x} - t \frac{s}{\sqrt{n}} = -0.40183 - 2.228 \times 0.00103 = -0.40412 \quad \text{and}$$

$$L_u = \bar{x} + t \frac{s}{\sqrt{n}} = -0.40183 + 2.228 \times 0.00103 = -0.39954$$

(b) Comparison among two or more means

The difference between two quantities X and Y to be measured is the quantity

$$m_{x-y} = m_x - m_y ,$$

and is estimated by $\bar{x} - \bar{y}$, where \bar{x} and \bar{y} are averages of a number of measurements of X and Y respectively.

Suppose that we are interested in knowing whether the difference m_{x-y} could be zero. This problem can be solved by the technique introduced in section 2, i.e., we could compute confidence limits for m_{x-y} , and if the upper and lower limits bracket zero, we could conclude that m_{x-y} may take the value zero; otherwise, we conclude that the evidence is against $m_{x-y} = 0$.

Let us assume that measurements of X and Y are independent with known variances σ_x^2 and σ_y^2 respectively.

By (2.10)

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} \quad \text{for } \bar{x} \text{ of } n \text{ measurements,}$$

$$\sigma_{\bar{y}}^2 = \frac{\sigma_y^2}{k} \quad \text{for } \bar{y} \text{ of } k \text{ measurements;}$$

then by (2.8)

$$\sigma_{\bar{x}-\bar{y}}^2 = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{k} .$$

Therefore the quantity,

$$z = \frac{(\bar{x} - \bar{y}) - 0}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{k}}} \quad (3.1)$$

is approximately normally distributed with mean zero and a standard deviation of one under the assumption $m_{x-y} = 0$.

If σ_x and σ_y are not known, but the two can be assumed to be approximately equal, e.g., \bar{x} and \bar{y} are measured by the same process, then s_x^2 and s_y^2 can be pooled by (2.15), or

$$s_p^2 = \frac{(n-1)s_x^2 + (k-1)s_y^2}{n+k-2}$$

This pooled computed variance estimates

$$\sigma^2 = \sigma_x^2 = \sigma_y^2, \text{ so that}$$

$$\sigma_{\bar{x}-\bar{y}}^2 = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{k} = \frac{n+k}{nk} \sigma^2.$$

Thus, the quantity

$$t = \frac{(\bar{x} - \bar{y}) - 0}{\sqrt{\frac{n+k}{nk}} \cdot s_p} \quad (3.2)$$

is distributed as Student's "t", and a confidence interval can be set about m_{x-y} with $\nu = n + k - 2$ and $P = 1 - \alpha$. If this interval does not bracket zero, we may conclude that the evidence is strongly against the hypothesis $m_x = m_y$.

As an example, we continue with the calibration of weights with NB' 10 gram. Data for the observed corrections during the months of September and October are listed in Table 6. It is desired to compare the means of observed corrections for the two sets of data in the two tables.

Here $n = k = 11$

$$\bar{x} = -.40183$$

$$\bar{y} = -.40454$$

$$s_x^2 = .000011669$$

$$s_y^2 = .000023813$$

$$s_p^2 = \frac{1}{2}(.000035482) = .000017741$$

$$\frac{n+k}{nk} = \frac{11 + 11}{121} = \frac{2}{11}$$

$$\sqrt{\frac{n+k}{nk}} s_p = \sqrt{\frac{2}{11}} \times .000017741 = .00180$$

For $1 - \alpha = .95$, $\nu = 20$, $t = 2.086$.

Therefore,

$$L_u = (\bar{x} - \bar{y}) + t\sqrt{\frac{n+k}{nk}} s_p = .00271 + 2.086 \times .00180$$

$$= .00646$$

$$L_l = (\bar{x} - \bar{y}) - t\sqrt{\frac{n+k}{nk}} s_p = -.00104$$

Since $L_l < 0 < L_u$, we conclude that there is no evidence against the hypothesis that the two observed average corrections are the same, or $m_x = m_y$. Note, however, that we would reach a conclusion of no difference wherever the magnitude of $\bar{x} - \bar{y}$ (.00271 mg.) is less than the half-width of the confidence interval ($2.086 \times .00180 = .00375$ mg.) calculated for the particular case. When the true difference m_{x-y} is large, the above situation is not likely to happen; but when the true difference is

Table 6

Computation of Confidence Limits for Observed Corrections,
NB' 10 gm.

<u>Date</u>	<u>Obs. Corr., mg.</u>
9-3-63	-.4020
9-24	-.4035
9-25	-.4051
9-25	-.3924
10-8	-.4070
10-9	-.4063
10-14	-.4028
10-16	-.4084
10-16	-.4039
10-18	-.4069
10-22	-.4116
	$\Sigma y_i = \frac{-4.4116}{-4.4499}$
	$\Sigma y_i^2 = 1.80038449$
	$\bar{y} = -.40454$
	$\frac{(\Sigma y_i)^2}{n} = 1.80014636$
	diff. = .00023813

$$s^2 = \frac{1}{n-1} \times .00023813 = .000023813$$

$s = .00488$ = computed standard deviation of an observed correction about the mean.

$$\frac{s}{\sqrt{n}} = .00147 = \text{computed standard deviation of the mean of eleven corrections}$$

$$t = 2.228 \text{ for } \alpha = .05 \text{ and } v = 10$$

$$L_l = -.40454 - .00328 = -.40782$$

$$L_u = -.40454 + .00328 = -.40126$$

small, say about .003 mg., then it is highly probable that a conclusion of no difference will still be reached. If a detection of difference of this magnitude is of interest, more measurements will be needed.

The following additional topics are treated in reference [1]:

- i. Sample sizes required under certain specified conditions - Tables A-8 and A-9.
- ii. σ_x^2 cannot be assumed to be equal to σ_y^2 - Section 3-3.1.2.
- iii. Comparison of several means by Studentized range - Sections 3-4, 15-4.

(c) Comparison of variances or ranges

As we have seen, the precision of a measurement process can be expressed in terms of the computed standard deviation, the variance, or the range. To compare the precision of two processes a and b, any of the three measures can be used, depending on the preference and convenience of the user.

Let s_a^2 be the estimate of σ_a^2 with ν_a degrees of freedom, and s_b^2 be the estimate of σ_b^2 with ν_b degrees of freedom. The ratio $F = s_a^2 / s_b^2$ has a distribution depending on ν_a and ν_b . Tables of upper percentage points of F are given in most statistical textbooks, e.g., ref. [1], table A-5 and section 4.2.

In the comparison of means, we were interested in finding out if the absolute difference between m_a and m_b could reasonably be zero; similarly, here we may be interested in whether $\sigma_a^2 = \sigma_b^2$, or $\sigma_a^2 / \sigma_b^2 = 1$. In practice, however, we are usually concerned with whether the imprecision of one process exceeds that of another process. We could, therefore, compute the ratio of s_a^2 to s_b^2 , and ask the question: If in fact $\sigma_a^2 = \sigma_b^2$, what is the probability of getting a value of the ratio as large as the one observed? For each pair of values of ν_a and ν_b , the tables list the values of F which are exceeded with probability α , the upper percentage point of the distribution of F . If the computed value of F exceeds this tabulated value of F_α , ν_a , ν_b , then we conclude that the evidence is against the hypothesis $\sigma_a^2 = \sigma_b^2$; if it is less, we conclude that σ_a^2 could be equal to σ_b^2 .

For example, we could compute the ratio of s_y^2 to s_x^2 given in section 3b immediately following Table 6. Here the degrees of freedom $\nu_y = \nu_x = 10$, the tabulated value of F which is exceeded 5% of the time for these degrees of freedom is 2.98, and

$$\frac{s_y^2}{s_x^2} = \frac{.000023813}{.000011669} = 2.041 .$$

Since 2.04 is less than 2.98, we conclude that there is no reason to believe that the precision of the calibration process in September and October is poorer than that of May. (Note that the observed correction of $-.3924$ on September 25 contributed to more than half of the total sum of squares. This point, no.63, falls outside of the control limits in Figure 5, and would have been cause for rejection of the calibrated values in practice.)

For small degrees of freedom the critical value of F is rather large, e.g., for $v_a = v_b = 3$, and $\alpha = .05$, the value of F is 9.28. It follows that a small difference between σ_a^2 and σ_b^2 is not likely to be detected with a small number of measurements from each process. The table below gives the approximate number of measurements required to have a four out of five chance of detecting whether σ_a is the indicated multiple of σ_b (while maintaining at .05 the probability of incorrectly concluding that $\sigma_a > \sigma_b$ when in fact $\sigma_a = \sigma_b$).

<u>Multiple</u>	<u>No. of Measurements</u>
1.5	39
2.0	15
2.5	9
3.0	7
3.5	6
4.0	5

Table A-11 in reference [1] gives the critical values of the ratios of ranges, and Tables A-20 and A-21 give confidence limits on the standard deviation of the process based on computed standard deviation.

4. Control Charts Technique for Maintaining Stability and Precision

A laboratory which performs routine measurement or calibration operations yields, as its daily product, numbers--averages, standard deviations, and ranges. The control chart techniques therefore could be applied to these numbers as products of a manufacturing process to furnish graphical evidence on whether the measurement process is in statistical control or out of statistical control. If it is out of control, these charts usually also indicate where and when the trouble occurred.

(a) Control Chart for Averages

The basic concept of a control chart is in accord with what has been discussed thus far. A measurement process with limiting mean m and standard deviation σ is

assumed. The sequence of numbers produced is divided into "rational" subgroups, e.g., by day, by a set of calibration, etc. The averages of these subgroups are computed. These averages will have a mean m and a standard deviation σ/\sqrt{n} where n is the number of measurements within each subgroup. These averages are approximately normally distributed.

In the construction of the control chart for averages, m is plotted as the center line, $m + k \frac{\sigma}{\sqrt{n}}$ and $m - k \frac{\sigma}{\sqrt{n}}$ are plotted as control limits, and the averages are plotted in an orderly sequence. If k is taken to be 3, we know that the chance of a plotted point falling outside of the limits, if the process is in control, is very small. Therefore, if a plotted point falls outside these limits, a warning is sounded and investigative action to locate the "assignable" cause that produced the departure, or corrective measures, are called for.

The above reasoning would be applicable to actual cases only if we have chosen the proper standard deviation σ . If the standard deviation is estimated by pooling the estimates computed from each subgroup and denoted by σ_w (within group), obviously differences, if any, between group averages have not been taken into consideration. When there are between-group differences, the variance of the individual \bar{x} is not $\frac{\sigma_w^2}{n}$, but, as we have seen before, $\sigma_b^2 + \frac{\sigma_w^2}{n}$, where σ_b^2 represents the variance due to differences between groups. If σ_b^2 is of any consequence as compared to σ_w^2 , many of the \bar{x} values would exceed the limits constructed by using σ_w alone.

Two alternatives are open to us--either remove the cause of the between-group variation; or, if such variation is a proper component of error, take it into account as was discussed under 2(d).

As an illustration of the use of a control chart on averages, we use again the NB'10 gram data. One hundred observed corrections for NB'10 are plotted in fig.5, including the two sets of data given under comparison of means (points 18 thru 28, and points 60 thru 71). A 3-sigma limit of 8.6 μg was used based on the "accepted" value of standard deviation.

We note that all the averages are within the control limits, excepting numbers 36, 47, 63, 85, and 87. Five in a hundred falling outside of the 3-sigma limits is more than predicted by the theory. No particular reasons, however, could be found for these departures.

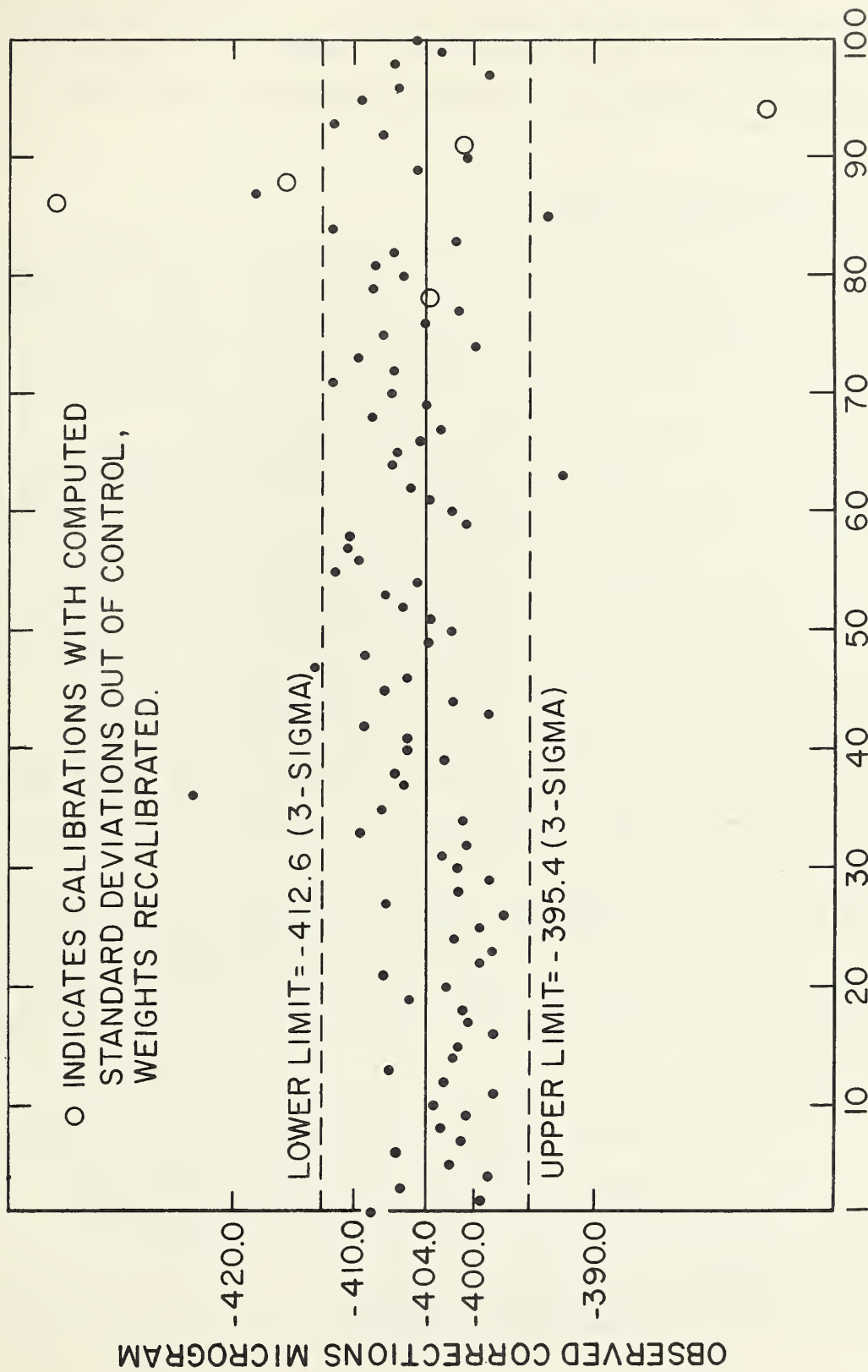


Figure 5 Control chart on \bar{x} for NB'10 gram.

Since the accepted value of the standard deviation was obtained by pooling a large number of computed standard deviations for within sets of calibrations, the graph indicates that a "between-set" component may be present. A slight shift upwards is also noted between the first 30 points and the remainder.

(b) Control Chart for Standard Deviations

The computed standard deviation, as we have seen, is a measure of imprecision. For a set of calibrations, however, the number of measurements is usually small, and consequently also the degrees of freedom. These computed standard deviations with few degrees of freedom can vary considerably by chance alone, even though the precision of the process remains unchanged. The control chart on the computed standard deviations (or ranges) is therefore an indispensable tool.

The distribution of s depends on the degrees of freedom associated with it, and is not symmetrical about m_s . The frequency curve of s is limited on the left side by zero, and has a long tail to the right. The limits, therefore, are not symmetrical about m_s . Furthermore, if the standard deviation of the process is known to be σ , m_s is not equal to σ , but is equal to $c_2\sigma$, where c_2 is a constant associated with the degrees of freedom in s .

The constants necessary for the construction of 3-sigma control limits for averages, computed standard deviations, and ranges, are given in most textbooks on quality control. Section 18-3 of ref. [1] gives such a table. A more comprehensive treatment on control chart is given in ASTM Manual on Quality Control of Materials, Special Technical Publication 15-C.

Unfortunately, the notation employed in quality control work differs in some respect from what is now standard in statistics, and correction factors have to be applied to some of these constants when the computed standard deviation is calculated by the definition given in this chapter. These corrections are explained in the footnote under the table.

As an example of the use of control charts on the precision of a calibration process, we will use data from NBS calibration of standard cells.^{1/} Standard cells in groups of four or six are usually compared with a NBS standard cell on ten separate days. A typical data sheet for a group of six cells, after all the

^{1/} Illustrative data supplied by Miss Catherine Law,
Electricity Division, National Bureau of Standards.

necessary corrections, appears in Table 7. The standard deviation of a comparison is calculated from the ten comparisons for each cell and the standard deviation for the average value of the ten comparisons is listed in the line marked SDA. These values were plotted as points 6 thru 11 in figure 6.

Let us assume that the precision of the calibration process remains the same. We can therefore pool the standard deviations computed for each cell (with nine degrees of freedom) over a number of cells and take this value as the current value of the standard deviation of a comparison, σ . The corresponding current value of standard deviation of the average of ten comparisons will be denoted by $\sigma' = \sigma/\sqrt{10}$. The control chart will be made on $s' = s/\sqrt{10}$.

For example, the SDA's for thirty-two cells calibrated between June 29 and August 8, 1962, are plotted as the first 32 points in figure 6. The pooled standard deviation of the average is 0.114 with 288 degrees of freedom. The between group component is assumed to be negligible.

Since $n = 10$, we find our constants for 3-sigma control limits on s' in section 18-3 of ref. [1] and apply the corrections as follows:

$$\text{Center line} = \sqrt{\frac{n}{n-1}} \quad c_2\sigma' = 1.111 \times .9227 \times .114 = .117$$

$$\text{Lower limit} = \sqrt{\frac{n}{n-1}} \quad B_1\sigma' = 1.111 \times .262 \times .114 = .033$$

$$\text{Upper limit} = \sqrt{\frac{n}{n-1}} \quad B_2\sigma' = 1.111 \times 1.584 \times .114 = .201$$

The control chart was constructed using these values of center line and control limits computed from the thirty-two calibrations. The standard deviations of the averages of subsequent calibrations are then plotted.

Three points in figure 6 far exceed the upper control limit. All three cells, which were from the same source, showed drifts during the period of calibration. A fourth point barely exceeded the limit. It is to be noted that the data here were selected to include these three points for purposes of illustration only, and do not represent the normal sequence of calibrations.

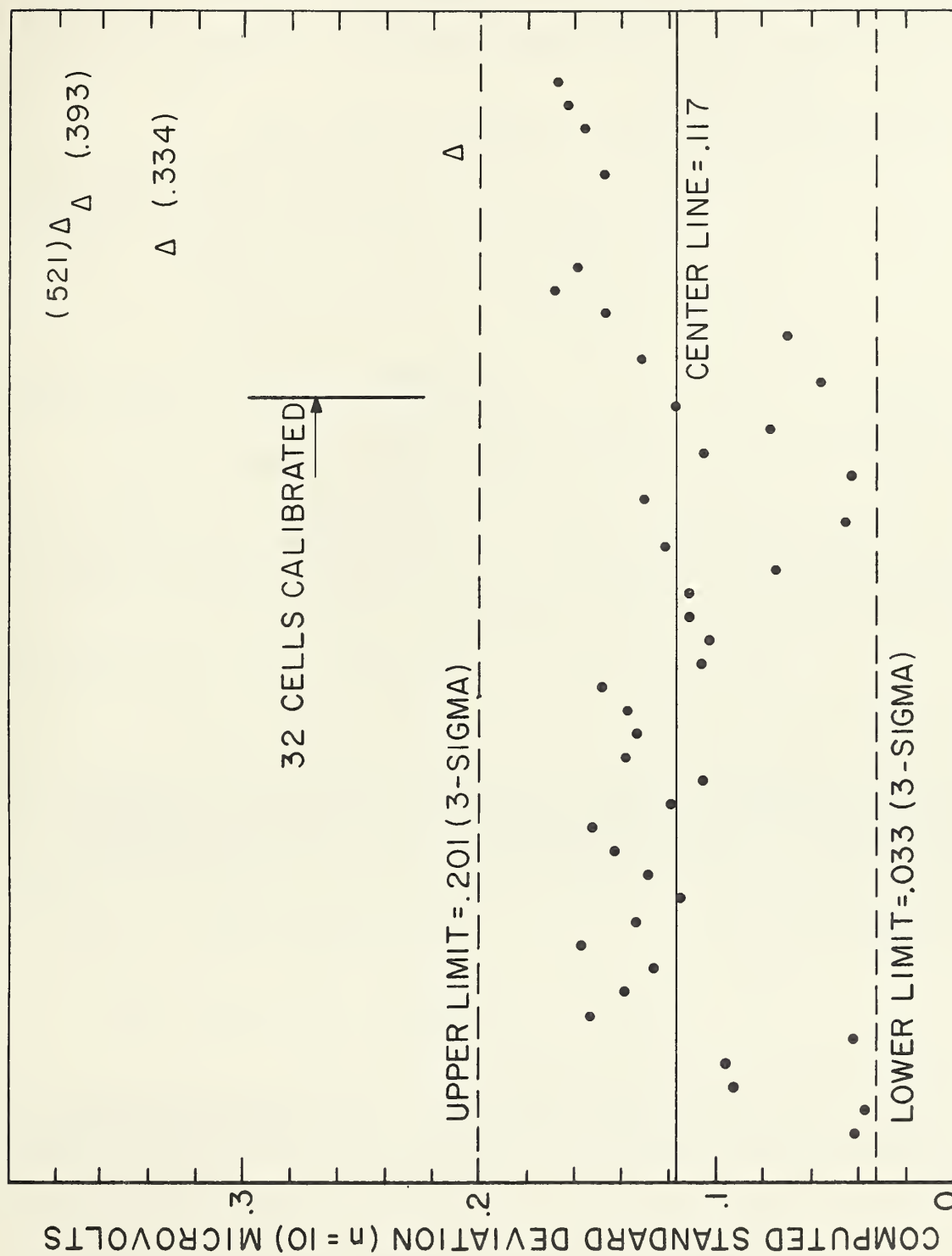
The main function of the chart is to justify the precision statement on the report of calibration, which is based on a value of σ estimated with perhaps thousands of degrees of freedom and which is shown to be in control. The report of calibration for these cells ($\sigma = .117 \pm .12$) could read:

Table 7

Calibration Data for Standard Cells

Day	Corrected EMF'S and Standard Deviations, Microvolts					
1	27.10	24.30	31.30	33.30	32.30	23.20
2	25.96	24.06	31.06	34.16	33.26	23.76
3	26.02	24.22	31.92	33.82	33.22	24.02
4	26.26	24.96	31.26	33.96	33.26	24.16
5	27.23	25.23	31.53	34.73	33.33	24.43
6	25.90	24.40	31.80	33.90	32.90	24.10
7	26.79	24.99	32.19	34.39	33.39	24.39
8	26.18	24.98	32.18	35.08	33.98	24.38
9	26.17	25.07	31.97	34.27	33.07	23.97
10	26.16	25.16	31.96	34.06	32.96	24.16
R	1.331	1.169	1.127	1.777	1.677	1.233
AVG	26.378	24.738	31.718	34.168	33.168	24.058
SD	0.482	0.439	0.402	0.495	0.425	0.366
SDA	0.153	0.139	0.127	0.157	0.134	0.116

Position	EMF (Volts)
1	1.0182264
2	1.0182247
3	1.0182317
4	1.0182342
5	1.0182332
6	1.0182240



CELL CALIBRATIONS

Figure 6 Control chart on s for the calibration of standard cells.

"Each value is the mean of 10 observations made between _____ and _____ .

Based on a standard deviation of 0.12 microvolts for the means, these values are correct to 0.36 microvolts relative to the volt as maintained by the national reference group."

5. Linear Relationship and Fitting of Constants by Least Squares

In using the arithmetic mean of n measurements as an estimate of the limiting mean, we have, knowingly or unknowingly, fitted a constant to the data by the method of least squares, i.e., we have selected a value \hat{m} for m such that

$$\sum_{i=1}^n (y_i - \hat{m})^2 = \sum_{i=1}^n d_i^2$$

is a minimum. The solution is $\hat{m} = \bar{y}$. The deviations $d_i = y_i - \hat{m} = y_i - \bar{y}$ are called residuals.

Here we can express our measurements in the form of a mathematical model

$$Y = m + \epsilon \quad (5.1)$$

where Y stands for the observed values, m the limiting mean (a constant), and ϵ the random error (normal) of measurement with a limiting mean zero and a standard deviation σ . By (2.1) and (2.9), it follows that

$$m_y = m + m_\epsilon = m,$$

and

$$\sigma_y^2 = \sigma^2.$$

The method of least squares requires us to use that estimator \hat{m} for m such that the sum of squares of the residuals is a minimum (among all possible estimators). As a corollary, the method also states that the sum of squares of residuals divided by the number of measurements n less the number of estimated constants p will give us an estimate of σ^2 , i.e.,

$$s^2 = \frac{\sum (y_i - \hat{m})^2}{n-p} = \frac{\sum (y_i - \bar{y})^2}{n-1}. \quad (5.2)$$

It is seen that the above agrees with our definition of s^2 .

Suppose Y , the quantity measured, exhibits a linear functional relationship with a variable which can be controlled or measured accurately; then a model can be written as

$$Y = a + bX + \epsilon, \quad (5.3)$$

where, as before, Y is the quantity measured, a (the intercept) and b (the slope) are two constants to be estimated, and ϵ the random error with limiting mean zero

and variance σ^2 . We set X at x_i , and observe y_i . For example, y_i might be the change in length of a gage block steel observed for n equally spaced temperatures x_i within a certain range. The quantity of interest is the coefficient of thermal expansion b .

For any estimates of a and b , say \hat{a} and \hat{b} , we can compute a value \hat{y}_i for each x_i , or

$$\hat{y}_i = \hat{a} + \hat{b} x_i.$$

If we require the sum of squares of the residuals

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

to be a minimum, then it can be shown that

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \text{and} \quad (5.4)$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x}. \quad (5.5)$$

The variance of Y can be estimated by

$$s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2} \quad (5.6)$$

with $n-2$ degrees of freedom since two constants have been estimated from the data.

The standard errors of \hat{b} and \hat{a} are respectively estimated by $s_{\hat{b}}$ and $s_{\hat{a}}$

where

$$s_{\hat{b}}^2 = \frac{s^2}{\sum (x_i - \bar{x})^2}, \quad (5.7)$$

$$s_{\hat{a}}^2 = s^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right]. \quad (5.8)$$

With these estimates and the degrees of freedom associated with s^2 , confidence limits can be computed for \hat{a} and \hat{b} for the confidence coefficient selected if we assume that errors are normally distributed.

Thus, the lower and upper limits of a and b respectively:

$$\begin{aligned}\hat{a} - t s_{\hat{a}} & \qquad \hat{a} + t s_{\hat{a}} , \\ \hat{b} - t s_{\hat{b}} & \qquad \hat{b} + t s_{\hat{b}} ,\end{aligned}$$

for the value of t corresponding to the degree of freedom and the selected confidence coefficient.

The following problems relating to a linear relationship between two variables are treated in reference [1] section 5-4.

- a. confidence intervals for a point on the fitted line,
- b. confidence band for the line as a whole,
- c. confidence interval for a single predicted value of Y
for a given x .

Polynomial and multivariate relationships are treated in Chapter 6 of the same reference.

6. References

In addition to the three general references given at the end of the last chapter, the following are selected with special emphasis on their ease of understanding and applicability in the measurement science.

Statistical Methods

- (1) Experimental Statistics, by M. G. Natrella, NBS Handbook 91, U.S. Government Printing Office, Washington 25, D.C., 1963 (504 pp.).
- (2) Statistical Methods for Chemists, by W. J. Youden, John Wiley and Sons, 1951 (126 pp.).
- (3) Statistical Method in Research and Production, by O. L. Davies, Hafner Publishing Co., New York, Third Edition, 1957 (396 pp.).

Textbooks

- (4) Introduction to Statistical Analysis, by W. J. Dixon and F. J. Massey, McGraw-Hill Co., 1957, Second Edition (488 pp.).
- (5) Statistical Theory and Methodology in Science and Engineering, by K.A. Brownlee, John Wiley and Sons, New York, 1960 (570 pp.).
- (6) Introduction to the Theory of Probability and Statistics by N. Arley and K. R. Buch, John Wiley and Sons, New York, 1950 (236 pp.).

